

Numerical differentiation with annihilators in noisy environment

Mamadou Mboup · Cédric Join · Michel
Fliess

September 3, 2008

Abstract Numerical differentiation in noisy environment is revised through an algebraic approach. For each given order, an explicit formula yielding a pointwise derivative estimation is derived, using elementary differential algebraic operations. These expressions are composed of iterated integrals of the noisy observation signal. We show in particular that the introduction of delayed estimates affords significant improvement. An implementation in terms of a classical finite impulse response (FIR) digital filter is given. Several simulation results are presented.

Keywords Numerical differentiation, Operational calculus, Orthogonal polynomials, Linear filtering

Mathematics Subject Classification (2000) 65D25, 44A40, 44A10

M. Mboup
UFR Mathématiques et Informatique, CRIP5, Université Paris Descartes,
45 rue des Saints-Pères, 75270 Paris cedex 06, France
Tel.: +33144553525, Fax: +33144553535
E-mail: mboup@math-info.univ-paris5.fr

C. Join
CRAN (CNRS-UMR 7039), Université Henri Poincaré (Nancy I), BP 239,
54506 Vandœuvre-lès-Nancy, France
E-mail: cedric.join@cran.uhp-nancy.fr

M. Fliess
Équipe MAX, LIX (CNRS-UMR 7161), École polytechnique,
91128 Palaiseau, France
E-mail: Michel.Fliess@polytechnique.edu

M. Mboup · C. Join · M. Fliess
EPI ALIEN INRIA, France.

1 Introduction

Numerical differentiation, *i.e.*, the derivative estimation of noisy time signals, is a longstanding difficult ill-posed problem in numerical analysis and in signal processing and control. It has attracted a lot of attention due to its importance in many fields of engineering and applied mathematics. A number of different approaches have been proposed. Methods based on observer design may be found in the control literature [7], [6], [18]. See also [19, 22, 34] for other approaches from the control literature. In signal processing, it is very common to cast the problem in terms of frequency domain digital filter design [31], [29], [5]. This is motivated by the observation that an ideal n^{th} order differentiator has a frequency response of magnitude ω^n . Another interesting approach, due to [2], consists in inverting the minimum phase transfer function of a properly designed numerical integrator to obtain a digital differentiator. Similar ideas have also been presented by [8], in the continuous-time context.

All these interesting approaches have been developed as candidate alternatives to the very classical one, based on least-squares polynomial fitting or (spline) interpolation. When completed with a regularization step, this classical approach may be very efficient (see *i.e.* [20], [9]) in off-line applications.

In this paper, the numerical differentiation problem is revised through the algebraic framework of parameter estimation initially presented in [15] (see also [14] and [25]). Our purpose is to improve a new approach which started in [11, 16], and in [13, 12, 10], for solving various questions in control and in signal and image processing. Given a smooth signal x and an order n , a key point of our approach is to consider $\frac{d^n x(t)}{dt^n} \big|_{t=\tau}$, for each fixed $\tau \geq 0$, as a single parameter to be estimated from a noisy observation of the signal. A pointwise derivative estimation therefore follows by varying τ . The general ideas leading to this estimation are exposed in section 2. As in the classical approaches, the starting point is an order $N \geq n$ truncation of the Taylor series expansion of the signal. One of the key feature of the proposed approach is to operate in the operational calculus domain [36], [27, 28]. Therein, an extensive use of differential elimination and a series of algebraic manipulations yield, back in the time domain, an explicit expression for the estimate of $\frac{d^n x(t)}{dt^n} \big|_{t=\tau}$ as an integral operator of the noisy observation within a short time interval $[\tau, \tau+T]$. The differential algebraic manipulations may be chosen in such a way that the time domain integral operators specialise to iterated integrals. This corresponds to a family of pointwise derivative estimators introduced in section 3. This section contains the main contributions of the paper. Therein, we establish a direct link in between our algebraic numerical differentiators with the well known fact that a scalar product of the signal $x(t)$ with a kernel having n vanishing moments essentially captures its n^{th} order derivative, provided x is smooth. Note however that drawing an efficient numerical differentiator from this general principle is another history, especially in a noisy setting. More precisely, recall that (see *e.g.* [3, Theorem 2.3.5, p. 31]): *If $f(t)$ belongs to a reproducing kernel Hilbert space \mathcal{H} , with reproducing kernel $\mathcal{K}_{\mathcal{H}}(\tau, t)$, analytic*

in both t and τ^* , then we have

$$\langle f(t), \frac{\partial^\ell}{\partial \tau^{*\ell}} \mathcal{K}_{\mathcal{H}}(\tau, t) \rangle = f^{(\ell)}(\tau). \quad (1)$$

We mention that, the model space \mathcal{H} and especially the evaluation point τ have to be judiciously chosen before the equation (1) yields some efficient derivative estimator. But, once such choices are met, high performance numerical differentiation method is obtained. Now, it is a happy fact that the proposed algebraic estimators inherently achieve these judicious choices. In particular, the estimators are as efficient as an appropriate time delay is admitted. We give an explicit expression for the corresponding delay. Section 4 tells us how to compute the estimates based on the samples of the noisy observation signal. For equally spaced samples, our numerical differentiators take the form of classical finite impulse response (FIR) digital filters. These filters are next used to present some simulation examples. To ease the presentation, all the proofs are deferred to an appendix.

2 Preliminaries

We consider the estimation of $x^{(n)}(t)$, the n^{th} order derivative of a smooth signal $x(t)$ defined on an interval $\mathcal{I} \subset \mathbb{R}_+$. In a concrete application, the signal $x(t)$ is not directly available but rather it is observed through a noise corruption. In all the sequel, $\varpi(t)$ will denote the noise and we set $y(t) = x(t) + \varpi(t)$ for the noisy observation. The problem is challenging since it is well known that differentiation tends to amplify the noise.

2.1 Operational domain

Let us ignore the noise for a moment. Assume that $x(t)$ is analytic on \mathcal{I} so that we may consider, without any loss of generality, the convergent Taylor expansion at $t = 0$

$$x(t) = \sum_{i \geq 0} x^{(i)}(0) \frac{t^i}{i!}. \quad (2)$$

With $N \geq n$, the truncated Taylor expansion

$$x_N(t) = \sum_{i=0}^N x^{(i)}(0) \frac{t^i}{i!}, \quad (3)$$

satisfies the differential equation $\frac{d^{N+1}}{dt^{N+1}} x_N(t) = 0$. In the operational domain, this reads as

$$s^{N+1} \hat{x}_N(s) = s^N x(0) + s^{N-1} \dot{x}(0) + \dots + s^{N-n} x^{(n)}(0) + \dots + x^{(N)}(0), \quad (4)$$

where $\hat{x}_N(s)$ stands for the operational analog of $x_N(t)$. In all the sequel, the operational analog of a signal $u(t)$ will be denoted as $\hat{u}(s)$ and, to ease the notation to follow, we will drop the argument s and write \hat{u} for short.

The basic step towards the estimation of $x^{(n)}(t), t \geq 0$ is the estimation of the coefficient $x^{(n)}(0)$ from the observation $y(t)$. Indeed, assume that an algorithm is already designed for this task. Then, for any $\tau \geq 0$, it is clear that applying the same algorithm on the new signal observation $y_\tau(t) \triangleq \text{Heaviside}(t)y(t + \tau)$, will yield an estimate for $x^{(n)}(\tau)$. We thus lay the focus on how to estimate $x^{(n)}(0)$ in (4).

2.1.1 Annihilators

Of course it is also possible to estimate all the coefficients $x^{(i)}(0), i = 0, \dots, N$ simultaneously. However, not only the coefficients $x^{(i)}(0), i \neq n$ are not necessary for the estimation of $x^{(n)}(t)$ as explained above, but also simultaneous estimation is more sensitive to noise and numerical computation errors [26]. These drawbacks are avoided in the proposed approach. All the terms $s^{N-i}x^{(i)}(0)$ in (4) with $i \neq n$, are consequently considered as undesired perturbations which we proceed to annihilate. For this, it suffices to find a linear differential operator, *i.e.*

$$\Pi = \sum_{\text{finite}} \varrho_\ell(s) \frac{d^\ell}{ds^\ell}, \quad \varrho_\ell(s) \in \mathbb{C}(s), \quad (5)$$

satisfying

$$\Pi \hat{x}_N = \varrho(s) x^{(n)}(0), \quad (6)$$

for some rational function $\varrho(s) \in \mathbb{C}(s)$. Such a linear differential operator, subsequently called an *annihilator* for $x^{(n)}(0)$, obviously exists and is not unique. It is also clear that to each annihilator Π , there is a unique $\varrho(s) \in \mathbb{C}(s)$ such that (6) holds. We shall say that Π and $\varrho(s)$ are *associated*.

2.1.2 Integral operators

A linear differential operator is said to be proper (resp. strictly proper) if, and only if, each $\varrho_\ell(s)$ in (5) is a proper (resp. strictly proper) rational function. We recall that a rational function $\frac{a}{b}$ is said to be proper (resp. strictly proper) if, and only if, $\deg a \leq \deg b$ (resp. $\deg a < \deg b$).

Now, if Π is an annihilator associated with $\varrho(s)$, then left multiplying both sides of (6) by a rational function $\sigma(s)$ immediately shows that $\sigma(s)\Pi$ is also an annihilator, associated with $\sigma(s)\varrho(s)$. It is therefore easy to see that

Lemma 1 *An annihilator Π and its associated rational function ϱ as in (6) may always be chosen strictly proper.*

Henceforth, we shall consider only strictly proper annihilators with strictly proper associated rational function. This means that only integral operators will be involved, once (6) is translated back in the time domain.

2.2 Time domain

We now replace the model x_N by the actual noisy observation y . With (5), equation (6) then becomes

$$\varrho(s)\tilde{x}_N^{(n)}(0) = \sum_{\text{finite}} \varrho_\ell(s) \frac{d^\ell}{ds^\ell} \hat{y}. \quad (7)$$

This equation provides us with an operational estimator of $x^{(n)}(0)$. It now remains to express the estimate $\tilde{x}_N^{(n)}(0)$, directly in the time domain. For this, remember that $\varrho(s)$ and $\varrho_\ell(s)$ are strictly proper rational (transfer) functions. Let $f(t)$ and $h_\ell(t)$ be their corresponding impulse responses and recall that, by the classical rules of operational calculus, $\frac{d}{ds}$ corresponds to multiplication by $-t$. Then equation (7) readily reads in the time domain as

$$f(t)\tilde{x}_N^{(n)}(0) = \sum_{\text{finite}} \int_0^t h_\ell(t-\alpha)(-1)^\ell \alpha^\ell y(\alpha) d\alpha. \quad (8)$$

Here, the variable t represents the estimation time. This equation has therefore to be considered for fixed t , say $t = T$. Now, for each estimation time $T > 0$ such that $f(T) \neq 0$, we obtain a numerical estimate $\tilde{x}_N^{(n)}(0)$. Examine this estimate. To do this, write $R_N(t) = x(t) - x_N(t)$ from (2) and (3) and decompose the observation y in $y(t) = x_N(t) + R_N(t) + \varpi(t)$. Rewriting (8) in the form

$$\begin{aligned} \tilde{x}_N^{(n)}(0) &= \int_0^T h(\alpha; T) \{x_N(\alpha) + R_N(\alpha) + \varpi(\alpha)\} d\alpha \\ &= x^{(n)}(0) + e_{R_N}(0) + e_{\varpi}(0), \end{aligned} \quad (9)$$

shows that the estimate is corrupted by two sources of error:

- the mismodelling, stemming from the truncation of the Taylor expansion, which contributes to $e_{R_N}(0)$ (the bias term) and
- the noisy environment which produces the error $e_{\varpi}(0)$ (the variance term).

Needless to say, the contribution of each one of these two sources may vary markedly from one estimator to another. So, the game of the next section will be to devise an adequate annihilator Π , along with its associated rational function ϱ . By adequateness, we mean the minimisation of the overall estimation error. However for N fixed, minimizing the mismodelling effect is, most often, in opposition with reducing the noise error contribution. As an example, note that the estimate $\tilde{x}_N^{(n)}(0)$ will vary with T in general. Now, for N fixed, choosing a smaller estimation time, T , will tend to reduce $|e_{R_N}(0)|$. But at the same time, a large T is required to filter out the noise showing that the choice for T should obey a compromise.

3 Pointwise derivative estimation

Now that we have fixed the notations and the guidelines of our methodology, we investigate in this section some detailed properties and performance of a class of pointwise derivative estimators. These will be derived from a particular family of annihilators. And, as we shall shortly see, the Jacobi orthogonal polynomials [35] are inherently connected with these estimators. A least squares interpretation then naturally follows [25], [26] and this leads to one of the main contribution of the paper: the numerical differentiation is as efficient as an appropriately chosen time delay is tolerated.

3.1 Purely integral estimators

A linear differential operator Π is said to be in *finite-integral* form if, and only if, each $\varrho_\ell(s)$ in (5) is in the form $\varrho_\ell(s) = \frac{1}{s} H_\ell(\frac{1}{s})$ for some polynomial H_ℓ . We consider the family of annihilators defined below.

Lemma 2 *For any $\kappa, \mu \in \mathbb{N}$, the differential operator*

$$\Pi_{\kappa, \mu}^{N, n} = \frac{1}{s^{N+\mu+1}} \frac{d^{n+\kappa}}{ds^{n+\kappa}} \cdot \frac{1}{s} \cdot \frac{d^{N-n}}{ds^{N-n}} \cdot s^{N+1} \quad (10)$$

is a finite-integral form annihilator for $x^{(n)}(0)$ and, it is associated with

$$\varphi_{\kappa, \mu, N}(s) = \frac{(-1)^{(n+\kappa)}(n+\kappa)!(N-n)!}{s^{\mu+\kappa+N+n+2}}. \quad (11)$$

This means that for each set of positive integers $\{\kappa, \mu, N\}$, with $N \geq n \geq 0$, the equation

$$\varphi_{\kappa, \mu, N}(s) \tilde{x}_N^{(n)}(0) = \Pi_{\kappa, \mu}^{N, n} \hat{y} \quad (12)$$

provides an operational estimator of the n^{th} order derivative of $x(t)$, observed by $y(t)$. We thus obtain a family of estimators parametrized by κ, μ and N . In order to emphasize the dependence on these parameters, we introduce the notation $x^{(n)}(0; \kappa, \mu; N)$ for $x_N^{(n)}(0)$. If $N = n$, the differential operator $\Pi_{\kappa, \mu}^{N, n}$ reduces to

$$\Pi_{\kappa, \mu}^n = \frac{1}{s^{n+\mu+1}} \frac{d^{n+\kappa}}{ds^{n+\kappa}} \cdot s^n. \quad (13)$$

Accordingly, the simplified notations $\tilde{x}^{(n)}(0; \kappa, \mu)$ and $\varphi_{\kappa, \mu}(s)$ will be used instead of $\tilde{x}^{(n)}(0; \kappa, \mu; n)$ and $\varphi_{\kappa, \mu, N}(s)$ respectively.

The next result will play a central rôle in the rest of the paper.

Theorem 1 *Let N, n, κ and μ be four positive integers, with $N \geq n$. Set $q = N - n$. Then the n^{th} order derivative estimate $\tilde{x}^{(n)}(0; \kappa, \mu; N)$, obtained from a Taylor expansion of order N , is uniquely expressible in the form*

$$\tilde{x}^{(n)}(0; \kappa, \mu; N) = \sum_{\ell=0}^q \lambda_\ell \tilde{x}^{(n)}(0; \kappa_\ell, \mu_\ell), \quad \lambda_\ell \in \mathbb{Q} \quad (14)$$

where $\kappa_\ell = \kappa + q + \ell$ and $\mu_\ell = \mu + \ell$. Moreover, if $q \leq n + \kappa$, then the coefficients λ_ℓ satisfy

$$\sum_{\ell=0}^q \lambda_\ell = 1.$$

This theorem shows that an n^{th} -order truncated Taylor expansion is appropriate for estimating the n^{th} -order derivative. Also, it characterises $\tilde{x}^{(n)}(0; \kappa, \mu; N)$ as a point in the \mathbb{Q} -affine hull of the set

$$\mathcal{S}_{\kappa, \mu, q} = \left\{ \tilde{x}^{(n)}(0; \kappa + q, \mu), \dots, \tilde{x}^{(n)}(0; \kappa, \mu + q) \right\}, \quad (15)$$

provided $q \leq n + \kappa$. Suppose now that, this affine hull is extended by allowing the barycentric coordinates λ_ℓ in (14) to be real rather than rational. By doing so, it is clear that any point in this new larger set will represent an n^{th} -order derivative estimate of $x(t)$ at the origin, in some meaningful sense. Characterising those points which minimise a given distance to $x^{(n)}(0)$ is an important question. This is the program of the following two subsections.

3.2 Minimizing the bias term

3.2.1 Jacobi orthogonal polynomials

We show in this subsection how the preceding special class of strictly proper differential operators are intimately connected to the Jacobi orthogonal polynomials. This connection, in turn, will allow us to attach a least squares interpretation to our estimators (see [25] for more details).

To begin let us recall that for a given signal \hat{u} , and a positive integer α , the time domain analog of $\hat{v} = \frac{1}{s^\alpha} \frac{d^\beta}{ds^\beta} \hat{u}$ is the iterated integral of order α , of $(-1)^\beta t^\beta u(t)$. Using the well known Cauchy formula for repeated integration, this reduces to a single integral

$$v(t) = \frac{1}{(\alpha - 1)!} \int_0^t (t - \tau)^{\alpha-1} (-1)^\beta \tau^\beta u(\tau) d\tau.$$

By a direct application of this rule of operational calculus and by recalling (13), (12) and (11), it becomes easy to see that, for any κ and μ ,

$$\tilde{x}^{(n)}(0; \kappa, \mu) = \frac{(\mu + \kappa + 2n + 1)!}{(\mu + n)!(\kappa + n)!} \int_0^T \frac{(T - \tau)^{\mu+n} \tau^{\kappa+n}}{T^{\mu+\kappa+2n+1}} y^{(n)}(\tau) d\tau. \quad (16)$$

Note that even though the noise function may not be differentiable, the above expression still makes sense because the derivation under the integral sign is only formal; it disappears upon integrating by parts. Nonetheless, to avoid any possible misunderstanding, we rewrite (16) by considering separately the

contribution due to the signal x and that due to the noise ϖ . For the former, we keep the form of (16) and for the latter, we integrate by parts, to obtain

$$\begin{aligned}\tilde{x}^{(n)}(0; \kappa, \mu) &= \gamma_{\kappa, \mu, n} \int_0^1 \tau^{\kappa+n} (1-\tau)^{\mu+n} x^{(n)}(T\tau) d\tau \\ &\quad + (-1)^n \frac{\gamma_{\kappa, \mu, n}}{T^n} \int_0^1 \frac{d^n}{d\tau^n} \{ \tau^{\kappa+n} (1-\tau)^{\mu+n} \} \varpi(T\tau) d\tau\end{aligned}\quad (17)$$

where $\gamma_{\kappa, \mu, n} \triangleq \frac{(\mu+\kappa+2n+1)!}{(\mu+n)!(\kappa+n)!}$. Here we have also used a change of variable to reduce the estimation interval $\mathcal{I}_0^T = [0, T]$ to $[0, 1]$.

Consider now $\{P_i^{\{\kappa, \mu\}}(t)\}_{i \geq 0}$, the Jacobi orthogonal polynomials (cf. [35, 1]) on the interval $[0, 1]$, associated to the weight function

$$w_{\kappa, \mu}(t) = t^{\kappa+n} (1-t)^{\mu+n}. \quad (18)$$

In all the sequel, $\|\cdot\|$ will denote the norm induced by the inner product defined, for two real valued functions f and g , by:

$$\langle f, g \rangle_{\{\kappa, \mu\}} = \int_0^1 f(t) w_{\kappa, \mu}(t) g(t) dt. \quad (19)$$

The short hand notation $\langle \cdot, \cdot \rangle$ will be used wherever there is no possible confusion. Upon noting that $P_0^{\{\kappa, \mu\}}(t) \equiv 1$ and

$$\|P_0^{\{\kappa, \mu\}}\|^2 = \frac{(\mu+n)!(\kappa+n)!}{(\mu+\kappa+2n+1)!} = \frac{1}{\gamma_{\kappa, \mu, n}}, \quad (20)$$

we have the following:

Proposition 1 *Let $x_{LS,1}^{(n)}(t)$ denote the first order least-squares polynomial approximation of $x^{(n)}(t)$ in the interval $[0, T]$. Then $\tilde{x}^{(n)}(0; \kappa, \mu)$ is given by*

$$\tilde{x}^{(n)}(0; \kappa, \mu) = x_{LS,1}^{(n)}(T\xi_1) + e_{\varpi}(0; \kappa, \mu), \quad (21)$$

where

$$\xi_1 = \frac{\kappa + n + 1}{\mu + \kappa + 2(n+1)} \quad (22)$$

is the root of $P_1^{\{\kappa, \mu\}}(t)$ and $e_{\varpi}(0; \kappa, \mu)$ is the noise contribution as given by the second term in the right hand side of (17).

3.2.2 Time-delayed derivative estimation

For a given $\tau \geq 0$, let us substitute, in all the preceding developments, $y(t)$ by $[\text{Heaviside}(\tau)y(\tau + t)]$. This simply amounts to moving the time origin from 0 to τ . As a result, we obtain $\tilde{x}^{(n)}(\tau; \kappa, \mu)$, the estimate of $x^{(n)}(\tau)$. Observe however that, the corresponding estimation is anti-causal, since the estimate at time τ is based on the signal observation during the time window $\mathcal{I}_{\tau+}^T = [\tau, \tau + T]$. Causal estimation is readily achieved by substituting $y(t)$ by $-\text{Heaviside}(t)y(\tau - t)$, for $\tau \geq T$. Doing so, we still obtain in $\tilde{x}^{(n)}(\tau; \kappa, \mu)$, the estimate of $x^{(n)}(\tau)$ but now from the signal observation on $\mathcal{I}_{\tau-}^T = [\tau - T, \tau]$.

These sliding windows thus induce a map $t \mapsto \tilde{x}^{(n)}(t; \kappa, \mu)$, as an image of the map $t \mapsto y(t)$ through all the preceding developments. The underlying operator will subsequently be called the (κ, μ) -algebraic numerical differentiator (AND) of order n . A (κ, μ) -AND will be termed *minimal* if it has been designed from the minimum truncation order of the Taylor expansion (3). It may be either causal or anti-causal, according to which estimation interval $\mathcal{I}_{\tau-}^T$ or $\mathcal{I}_{\tau+}^T$ is used.

Now, it is straightforward to infer from Proposition 1 that, the noise contribution $e_{\varpi}(t; \kappa, \mu)$ apart,

$$\tilde{x}^{(n)}(t; \kappa, \mu) = x_{LS,1}^{(n)}(t - T\xi_1), \quad t \geq T,$$

for the causal variant. As a consequence, we have:

Corollary 1 *Let $x(t)$ be observed through a noise corruption $\varpi(t)$ by $y(t) = x(t) + \varpi(t)$. The causal n^{th} order minimal (κ, μ) -algebraic numerical differentiation of $y(t)$ is time-delayed:*

$$\tilde{x}^{(n)}(t; \kappa, \mu) \approx x^{(n)}(t - \tau_1), \quad t \geq T. \quad (23)$$

The delay τ_1 reads as $\tau_1 = T\xi_1$ where ξ_1 is given by (22).

Similarly, we obviously have for the anti-causal variant,

$$\tilde{x}^{(n)}(t; \kappa, \mu) \approx x^{(n)}(t + \tau_1), \quad t \geq 0. \quad (24)$$

Clearly, the above approximations become exact for polynomial signals of degree not exceeding n , in the noise free context. But since ξ_1 , in Proposition 1, is the root of $P_1^{\{\kappa, \mu\}}(t)$, we deduce:

Corollary 2 *The approximations in (23) and (24) are exact for polynomial signals of degree up to $n + 1$.*

3.2.3 Delay and model complexity

In this paragraph, we investigate the non minimality effects on the algebraic numerical differentiators. We shall first establish that non minimality allows to compensate the previous delay. This is important because a delay may not

be tolerable in some real time applications¹. We shall also show that for a fixed modelling order N , the absence of delay undergoes a performance loss. Finally, the question of the “best” choice of the barycentric coordinates in (14) will be approached.

To begin, let us recall some well-known facts. Consider the subspace of $L_2([0, 1])$, defined by

$$\mathcal{H}_q = \text{span}\{P_0^{\{\kappa, \mu\}}(t), P_1^{\{\kappa, \mu\}}(t), \dots, P_q^{\{\kappa, \mu\}}(t)\}. \quad (25)$$

Equipped with the inner product (19), \mathcal{H}_q is clearly a reproducing kernel Hilbert space [4], [3], [32], with the reproducing kernel

$$\mathcal{K}_q(\tau, t) = \sum_{i=0}^q \frac{P_i^{\{\kappa, \mu\}}(\tau) P_i^{\{\kappa, \mu\}}(t)}{\|P_i^{\{\kappa, \mu\}}\|^2}. \quad (26)$$

The reproducing property implies that for any function f defined on $[0, 1]$, we have

$$\langle \mathcal{K}_q(\tau, \cdot), f(\cdot) \rangle = f_q(\tau),$$

where f_q stands for the orthogonal projection of f on \mathcal{H}_q . In the rest of the paper, we will set $q = N - n$ and assume that $q \geq 0$.

Theorem 2 Assume that $q \leq \kappa + n$ and let

$$x_{LS,q}^{(n)}(t) \triangleq \sum_{i=0}^q \frac{\langle P_i^{\{\kappa, \mu\}}(\tau), x^{(n)}(T\tau) \rangle}{\|P_i^{\{\kappa, \mu\}}\|^2} P_i^{\{\kappa, \mu\}}(t/T), \quad (27)$$

denote the least-squares q^{th} order polynomial approximation of $x^{(n)}(\cdot)$ in the interval $\mathcal{I}_{0-}^T = [0, T]$. Then $\tilde{x}^{(n)}(0; \kappa, \mu; N)$ is given by

$$\tilde{x}^{(n)}(0; \kappa, \mu; N) = x_{LS,q}^{(n)}(0) + e_{\varpi}(0; \kappa, \mu; N), \quad (28)$$

where $e_{\varpi}(0; \kappa, \mu; N)$ represents the corresponding noise contribution.

Remark 1 This is for the anti-causal variant. To obtain the causal counterpart, just replace $y(t)$ for $t \in [0, T]$ by $-y(T - t)$.

We immediately deduce from the above theorem that the corresponding causal (κ, μ) -AND is delay-free:

$$\tilde{x}^{(n)}(t; \kappa, \mu; N) \approx x^{(n)}(t), \quad t \geq T. \quad (29)$$

Now, since $x_{LS,q}^{(n)}(Tt)$ is the orthogonal projection of $x^{(n)}(Tt)$, $t \in [0, 1]$ on \mathcal{H}_q , we may rewrite (28) as

$$\begin{aligned} \tilde{x}^{(n)}(0; \kappa, \mu; N) &= \langle \mathcal{K}_q(0, t), x^{(n)}(Tt) \rangle + e_{\varpi}(0; \kappa, \mu; N) \\ &= \sum_{\ell=0}^q \lambda_{\ell} \tilde{x}^{(n)}(0; \kappa + q - \ell, \mu + \ell), \end{aligned} \quad (30)$$

¹ For instance, it is known that introducing delayed signals in a control loop of a system tends to destabilize it [30], [33].

where we recall (14) (for $q \leq \kappa + n$). As already suggested, we henceforth allow the barycentric coordinates $\lambda_\ell, \ell = 0, \dots, q$ to belong to \mathbb{R} instead of being restricted in \mathbb{Q} . We obtain the next result.

Proposition 2 *Let κ, μ, n and q be four positive integers as above. For any $\xi \in [0, 1]$, there exists a unique set of real coordinates $\lambda_\ell(\xi) \in \mathbb{R}, \ell = 0, \dots, q$ such that*

$$\sum_{\ell=0}^q \lambda_\ell(\xi) \tilde{x}^{(n)}(0; \kappa_\ell, \mu_\ell) = \langle \mathcal{K}_q(\xi, t), x^{(n)}(Tt) \rangle + e_{\varpi}^\xi(0; \kappa, \mu; n + q), \quad (31)$$

for some noise contribution $e_{\varpi}^\xi(0; \kappa, \mu; n + q)$.

Moreover, these coordinates satisfy

$$\sum_{\ell=0}^q \lambda_\ell(\xi) = 1$$

and, unless all the $\lambda_\ell(\cdot)$ are equal to zero except one, the following holds:

$$\min_{\ell} \lambda_\ell(\cdot) < 0.$$

For each $\xi \in [0, 1]$, this proposition establishes a correspondance between $x_{LS,q}^{(n)}(T\xi)$ and a point in the \mathbb{R} -affine hull of $\mathcal{S}_{\kappa,\mu,q}$ in (15).

Let $\tilde{x}_\xi^{(n)}(0; \kappa, \mu; N)$ denotes such a point as given in (31) and consider as before, the map $t \mapsto \tilde{x}_\xi^{(n)}(t; \kappa, \mu; N)$:

$$\tilde{x}_\xi^{(n)}(0; \kappa, \mu; N) = \langle \mathcal{K}_q(\xi, t), x^{(n)}(Tt) \rangle + e_{\varpi}^\xi(0; \kappa, \mu; n + q). \quad (32)$$

Then, following the previous paragraph we clearly see that the corresponding causal algebraic numerical differentiator is time delayed:

$$\tilde{x}_\xi^{(n)}(t; \kappa, \mu; N) \approx x^{(n)}(t - \tau), \quad t \geq T, \quad (33)$$

where the delay is given by $\tau = T\xi$.

In order to simplify some notations to come, we turn to the anti-causal variant (replace $x^{(n)}(t - \tau), t \geq T$ above by $x^{(n)}(t + \tau), t \geq 0$). Let now $0 < \xi_1^{\{q+1\}} < \xi_2^{\{q+1\}} < \dots < \xi_{q+1}^{\{q+1\}} < 1$ represent the $(q + 1)$ zeros of $P_{q+1}^{\{\kappa, \mu\}}$. Then, the noise aside, one may readily check that for each $i = 1, \dots, q + 1$,

$$\begin{aligned} \tilde{x}_{\xi_i^{\{q+1\}}}^{(n)}(t; \kappa, \mu; n + q) &= \langle \mathcal{K}_{q+1}(\xi_i^{\{q+1\}}, u), x^{(n)}(t + Tu) \rangle \\ &= x_{LS,q+1}^{(n)}(T\xi_i^{\{q+1\}}). \end{aligned} \quad (34)$$

In other words, in the noise-free case, the approximation (33) with $\xi = \xi_i^{\{q+1\}}, 1 \leq i \leq q + 1$ would be exact for all polynomial signals up to degree $N + 1$ although this approximation is based on a Taylor expansion of order N . This has an important implication as shown next.

3.2.4 Delay and precision

To continue the discussion, we establish in Proposition 3 below that the quality of the estimation may be significantly improved by admitting a delay. When relating this result with the last remark of the preceding subsection, we somehow recover the paradigm of model section [24]: *a simple and approximate model may outperform a precise and more complex one.*

To begin, let us expend $x^{(n)}(t + Tu)$, for each fixed t , as

$$x^{(n)}(t + Tu) = \sum a_i(t) P_i^{\{\kappa, \mu\}}(u), \quad u \in [0, 1], \quad (35)$$

and set $\xi_i^{\{q\}}, i = 1, \dots, q$ for the zeros of $P_q^{\{\kappa, \mu\}}$, in increasing order. Suppose that $q = N - n \geq 1$. For some given i , consider first $\tilde{x}_{\xi_i^{\{q\}}}^{(n)}(t; \kappa, \mu; N - 1)$, the time-delayed $\tau = T\xi_i^{\{q\}}$ estimate (34), based on an $(N - 1)$ -order truncated Taylor expansion. Using (35) and (34), the estimation error due to the truncation reads, at each time t , as:

$$e_{R_{N-1}}^{\xi_i^{\{q\}}}(t; \kappa, \mu; n + q - 1) = a_{q+1}(t) P_{q+1}^{\{\kappa, \mu\}}(\xi_i^{\{q\}}) + \sum_{i \geq 2} a_{q+i}(t) P_{q+i}^{\{\kappa, \mu\}}(\xi_i^{\{q\}}). \quad (36)$$

Next, suppose that 1) we do not want the estimation to be time-delayed and 2) we have a more precise model, namely a Taylor truncation of order N instead of $N - 1$. Accordingly, the estimation error $e_{R_N}(t; \kappa, \mu; n + q)$ corresponding to the delay-free estimate $\tilde{x}^{(n)}(t; \kappa, \mu; N)$ in (28)-(30) and stemming from the mismodelling is given by:

$$e_{R_N}(t; \kappa, \mu; n + q) = a_{q+1}(t) P_{q+1}^{\{\kappa, \mu\}}(0) + \sum_{i \geq 2} a_{q+i}(t) P_{q+i}^{\{\kappa, \mu\}}(0). \quad (37)$$

Let us now compare the magnitude of these two errors. For that, let us notice that for T small enough, it is reasonable to approximate the above error by the first term of the expansion (37):

$$e_{R_{N-1}}(t; \kappa, \mu; n + q) \approx a_{q+1}(t) P_{q+1}^{\{\kappa, \mu\}}(0) \quad (38)$$

and likewise for $e_{R_{N-1}}^{\xi_i^{\{q\}}}(t; \kappa, \mu; n + q - 1)$ in (36). Using the three terms recurrence relation of the Jacobi polynomials and the recurrence relation (61) (see proof of Theorem 2), we obtain the following:

Proposition 3 *Denote by $z_q(\kappa, \mu)$ and $Z_q(\kappa, \mu)$ the lower and upper bounds of the smallest and largest zeros of $P_q^{\{\kappa, \mu\}}$, respectively. Then*

$$\frac{\left| e_{R_{N-1}}^{\xi_1^{\{q\}}}(t; \kappa, \mu; n + q - 1) \right|}{|e_{R_N}(t; \kappa, \mu; n + q)|} < \theta_q(\kappa, \mu) \left(1 - \frac{z_q(\kappa, \mu)}{Z_{q-1}(\kappa, \mu)} \right)^{q-1}, \quad (39)$$

where

$$\theta_q(\kappa, \mu) = \frac{q(q + n + \mu)(2(q + n) + \mu + \kappa + 2)}{(q + n + \kappa + 1)(q + 2n + \mu + \kappa + 1)(2q + 2n + \mu + \kappa)} < 1. \quad (40)$$

Explicit expressions for $z_q(\kappa, \mu)$ and $Z_q(\kappa, \mu)$ may be obtained from [21, Theorem 2]. Anyway, for $i = 1$ in (36), which amounts to choosing the minimum delay² $\tau = T\xi_1^{\{q\}}$, a significant improvement is achieved over a delay-free setting, even with a more precise model.

Remark 2 For the causal variant, the minimum delay will be given by $\tau = T\xi_q^{\{q\}}$ and all the preceding arguments remain valid upon interchanging κ with μ , and τ with $1 - \tau$.

Finally, we have seen that a causal non minimal (κ, μ) -AND may provide either delay-free or time-delayed estimation. The performance obtained in the delay-free case are worst than for the time-delayed counterpart, even though a more precise signal model is used. Admitting a delay significantly improves the quality of an (κ, μ) -AND. Thus, we recover the paradigm of model selection [24]. On the other hand, it appears that the question of a “good” choice for the barycentric coordinates reduces to the choice of the time-delay, provided the noise effect is ignored.

3.3 Reducing the noise effect

In the preceding analysis, we have confined our attention to minimizing the mismodelling error e_{R_N} . Here, we investigate the effect of the noise. The analysis relies on the very classical properties of linear time-invariant filtering. Basically, it amounts to adjust the free parameters of a filter in order to maximize the output signal-to-noise ratio (SNR). Clearly, this approach is in contrast with the previous local analysis.

3.3.1 Implementation issues

As a matter of fact, a causal (κ, μ) -AND is equivalent to a stable and causal linear time-invariant filter, with finite impulse response. To see this, rewrite the equation (17) for the causal variant,

$$\tilde{x}^{(n)}(t; \kappa, \mu) = \frac{(-1)^n \gamma_{\kappa, \mu, n}}{T^n} \int_0^1 \frac{d^n}{d\tau^n} \{ \tau^{\kappa+n} (1 - \tau)^{\mu+n} \} y(t - T\tau) d\tau, \quad (41)$$

where t stands for the present time. Setting $y_T(\tau) = y(T\tau) = x(T\tau) + \varpi(T\tau)$ and

$$h_{\kappa, \mu}(\tau) \triangleq \frac{(-1)^n \gamma_{\kappa, \mu, n}}{T^n} \text{rect}(\tau) \frac{d^n}{d\tau^n} w_{\{\kappa, \mu\}}(\tau), \quad (42)$$

where $\text{rect}(\cdot)$ is the characteristic function of the interval $[0, 1]$, one may immediately rewrite (41) as the convolution product:

$$\tilde{x}^{(n)}(Tt; \kappa, \mu) = \int_0^\infty h_{\kappa, \mu}(\tau) y_T(t - \tau) d\tau. \quad (43)$$

² We still use the term *delay* although one should rather talk about *advance* as long as the anti-causal variant is in concern.

3.3.2 Minimizing the output noise variance

Given κ, μ and n , we consider that the estimation time T is chosen so that the mismodelling error is small. We are now interested in describing the points in the \mathbb{R} -affine hull of the set $\mathcal{S}_{\kappa, \mu, q}(t) = \{\tilde{x}^{(n)}(t; \kappa + q - \ell, \mu - \ell), \ell = 0, \dots, q\}$ which minimise the noise effect. To proceed, we suppose that the noise is a wide-sense stationary random process. Consider

$$\tilde{x}^{(n)}(Tt; \kappa, \mu; N) = \int_0^\infty \sum_{\ell=0}^q \lambda_\ell h_{\kappa+q-\ell, \mu+\ell}(\tau) y_T(t-\tau) d\tau, \quad (44)$$

a point, with barycentric coordinates $\lambda_\ell, \ell = 0, \dots, q$, in this affine hull. Set $\boldsymbol{\lambda} = [\lambda_0 \dots \lambda_q]^t$ where the superscript t denotes transposition. Then one may readily check that the associated output noise variance is

$$\begin{aligned} \text{var}(e_\varpi) &= \sum_{\ell, m}^q \lambda_\ell \lambda_m \int_0^1 \int_0^1 h_{\kappa_\ell, \mu_\ell}(\tau) \mathcal{C}(T(t-\tau)) h_{\kappa_m, \mu_m}(\tau) d\tau dt, \\ &= \boldsymbol{\lambda}^t \mathcal{R} \boldsymbol{\lambda}, \end{aligned} \quad (45)$$

where $\kappa_i = \kappa + q - i$, $\mu_i = \mu + i$, $\mathcal{C}(\tau) \triangleq \mathbb{E}[\varpi(t)\varpi(t-\tau)]$ stands for the noise covariance function and \mathcal{R} is the $(q+1)$ -square matrix with entries

$$\mathcal{R}_{i,j} = \int_0^1 \int_0^1 h_{\kappa+q-i, \mu+i}(\tau) \mathcal{C}(T(t-\tau)) h_{\kappa+q-j, \mu+j}(\tau) d\tau dt.$$

Define by \mathbf{u} the $(q+1)$ -dimensional real vector, with 1 in each entry. Now, the barycentric coordinates $\boldsymbol{\lambda}_\varpi$ minimizing this variance are immediately given by

$$\boldsymbol{\lambda}_\varpi = \frac{\mathcal{R}^{-1} \mathbf{u}}{\mathbf{u}^t \mathcal{R}^{-1} \mathbf{u}}. \quad (46)$$

By proposition 2, we know that $\tilde{x}^{(n)}(t; \kappa, \mu; N)$ associated to $\boldsymbol{\lambda}_\varpi$ is in the form

$$\tilde{x}^{(n)}(t; \kappa, \mu; N) = \langle \mathcal{K}_q(\xi, \tau), (-1)^n x^{(n)}(t - T\tau) \rangle + e_\varpi^\xi(t; \kappa, \mu; N),$$

for some $\xi \in [0, 1]$ if, and only if, at least one entry of $\boldsymbol{\lambda}_\varpi$ is strictly negative. However, the barycentric coordinates which minimise the noise variance (45) do not fulfill this requirement, in general. If $\varpi(t)$ is a (bandlimited) white noise for instance, then we directly verify that (46) yields

$$\boldsymbol{\lambda}_\varpi = \frac{1}{q+1} \mathbf{u}, \quad (47)$$

showing that the minimum output mean square error is achieved by the centroid of the points $\tilde{x}^{(n)}(t; \kappa + q - \ell, \mu - \ell), \ell = 0, \dots, q$.

If, more generally, we consider the (κ, μ) -AND associated now to the points in the convex hull of $\mathcal{S}_{\kappa, \mu, q}(t)$, then we obtain a class of differentiators different from the one described so far. The difference is, as stated by proposition 2, that these (κ, μ) -AND do not have the orthogonal projection principle attached to them.

We postpone the detailed analysis of this class to a forthcoming paper and go directly to the simulation studies.

4 Numerical experiments

4.1 A digital FIR filter

Before all, let us remind that the numerical estimates are obtained from (44) which we rewrite as

$$\tilde{x}^{(n)}(Tt; \kappa, \mu; N) = \int_0^1 g(\tau) y_T(t - \tau) d\tau, \quad (48)$$

with the simplified notation $g(t) = \sum_{\ell=0}^q \lambda_\ell h_{\kappa+q-\ell, \mu+\ell}(\tau)$. This is a very classical input-output linear time-invariant filtering relation, the implementation of which is easy. To further simplify the implementation, we consider that only discrete-time samples of the observation y are available. We assume that the samples are regularly spaced³, with a sampling period T_s and we use the common notation $y_m = y(mT_s)$. For an estimation time $T = MT_s$, we set $g_m = g(mT_s/T)$, $m = 0, \dots, M$. These numbers g_m are computed once for all, after all the parameters $\lambda_\ell, \kappa, \mu, n, q$ have been selected.

Let now W_m and $t_m = mT_s/T$, $m = 0, \dots, M$ be the weights and abscissas associated with a given numerical integration method as in

$$\int_0^1 f(t) dt \approx \sum_{m=0}^M W_m f(t_m).$$

For each sampling time $t_\ell = \ell T_s$, we compute

$$\tilde{x}^{(n)}(t_\ell; \kappa, \mu; N) \approx \tilde{x}_\ell^{(n)} = \sum_{m=0}^M W_m g_m y_{\ell-m}. \quad (49)$$

The n^{th} order derivative estimate is thus obtained from the output of a classical FIR digital filter with impulse response $c_m = W_m g_m$, $m = 0, \dots, M$, where the input is the noisy observation samples signal. In all the subsequent simulation, we use the work horse trapezoidal method, corresponding to

$$W_0 = W_M = \frac{T_s}{2}, \quad \text{and} \quad W_m = T_s, \quad m = 1, \dots, M-1.$$

4.2 Simulation results

Let $y(t) = x(t) + \varpi(t)$, $0 \leq t \leq 5$ (without unit), be the noisy measurement of the signal

$$x(t) = \tanh(t-1) + e^{-t/1.2} \sin(6t + \pi).$$

We consider a sampling period (given without unit) of $T_s = 1/200$ which corresponds to 1000 samples in the interval $[0, 5]$. The samples of the noise are simulated from a zero-mean white Gaussian *iid* sequence. The variance

³ Observe that non uniform sampling can also be handled in exactly the same way.

is adjusted in such a way that the signal to noise ratio in dB⁴, *i.e.*, $SNR = 10 \log_{10} \left(\frac{\sum |y_i|^2}{\sum |\varpi_i|^2} \right)$, corresponds to $SNR = 25$ dB (see figure 1).

Begin with the first order derivative. The estimates $\hat{x}^{(1)}(t; 0, 0)$, obtained with the minimal (0,0)-AND are displayed in figure 2 (dotted line). Each estimation window $\mathcal{I}_{t-}^T = [t - T, t]$ has a width $T = 60T_s$. We also depict (dashed line) the exact derivative of the noise-free signal in order to gauge the estimation accuracy. Compare with the solid line curve, obtained by a left shift of the dotted line curve with an amount corresponding to the delay $\xi_1 T$ in equation (22). With $N = 1$, the linear signal model results in poor estimates on the intervals where the signal's dynamic is strong. For high signal-to-noise ratio, the estimates may be improved by reducing the estimation time T .

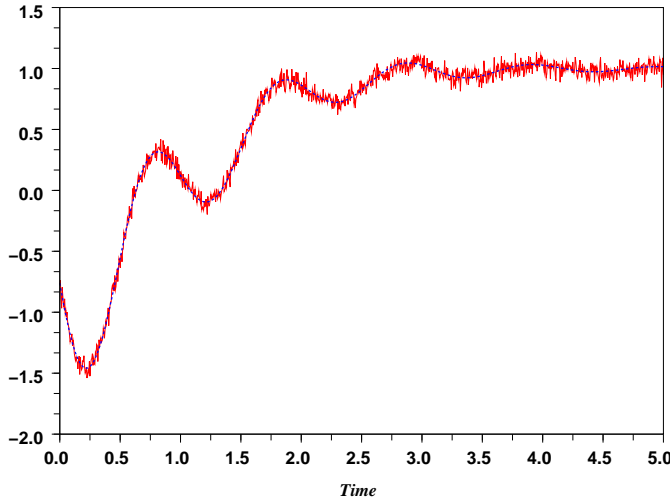


Fig. 1 Noisy observation signal, $SNR = 25dB$.

Alternatively, one may consider a richer signal model *e.g.* with $N = 2$.

This is the case of the next experiment. We consider the non minimal estimator from (10)-(12), with sliding windows of width $T = 110T_s$. The estimates $\hat{x}^{(1)}(t; 0, 0; 2)$ are plotted (solid line) in figure 3 below. There is no estimation delay, as expected. However, the performance significantly degrades as compared to the preceding results although the signal model is more precise. If we now relax the delay-free constraint, it becomes possible to take advantage of the more flexible second order model for the signal. This is illustrated in the following simulation (see Figure 4), where we keep the same settings for T , but with $\kappa = \mu = 1$ now. The solid line curve in Figure 4 represents the estimates obtained as in Proposition 2, with the barycentric coordinates given by equation (68) where $\xi = \xi_2^{\{2\}}$ is the largest root of $P_2^{\{\kappa, \mu\}}$. For comparison

⁴ See, *e.g.*, [17] for this well known concept in signal processing

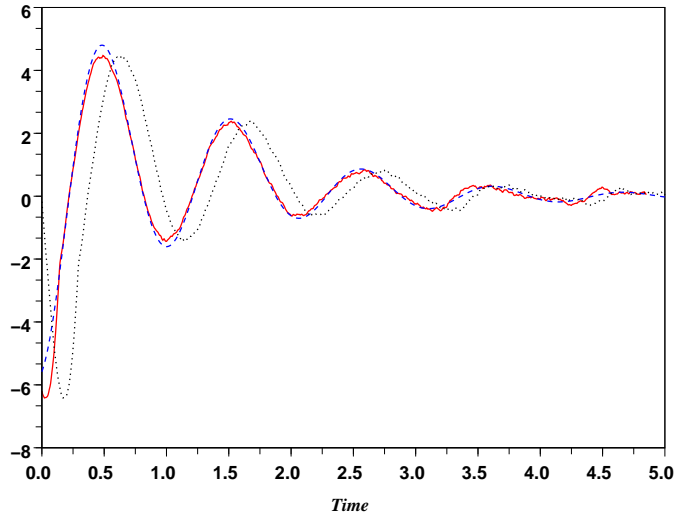


Fig. 2 Estimation of the signal derivative: minimal (0,0)-AND.

purpose, the curve has been shifted by a number of samples corresponding to the delay $\xi_2^{\{2\}}T$. The same type of behaviors are also observed for the esti-

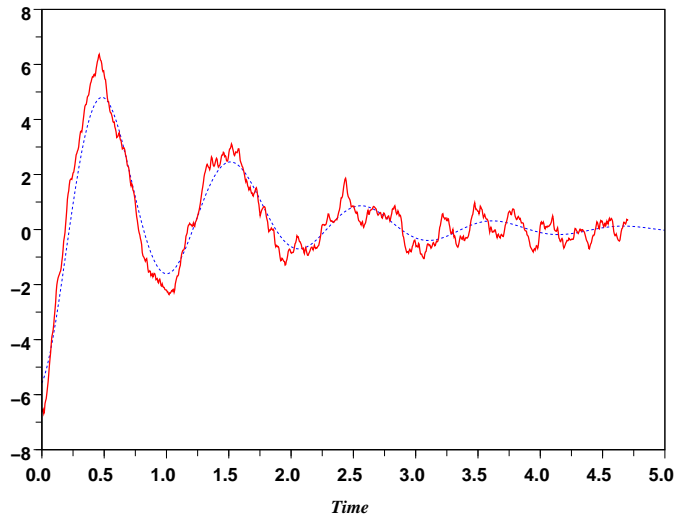


Fig. 3 Non minimal, delay-free derivative estimation.

mation of the higher order derivatives. As an illustration, we give in figure 5 and figure 6 below the estimation results corresponding to the second order derivative from the noisy observation depicted in figure 1.

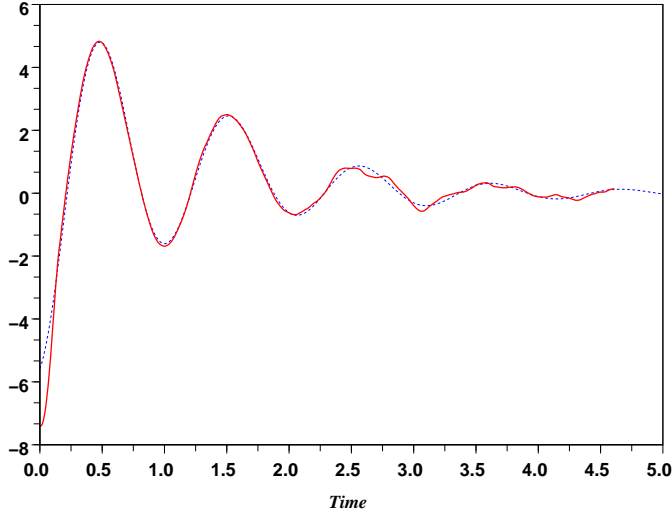


Fig. 4 Non minimal, delayed derivative estimation.

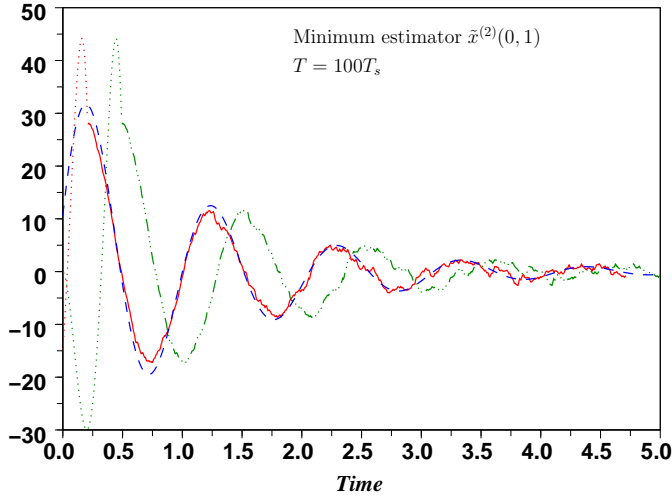


Fig. 5 Estimation of the second order derivative: minimal (0,1)-AND.

We represent the transient components of the corresponding filter's output by the dotted parts of the curves. The dash-dotted line curves are the delayed estimates. The predicted values of the respective delays match the experiment as can be seen by comparing the exact second order derivative (dashed line curves) with the shifted versions of the estimates (solid line curves).

Observe that for $q = 1$ and for all n , the filter corresponding to the minimal (κ, μ) -AND coincides with that which minimises the output noise variance for the same κ and μ . More generally, we may verify that any minimal (κ, μ) -AND, with small $|\kappa - \mu|$, is close to the corresponding minimum variance (κ, μ) -AND,

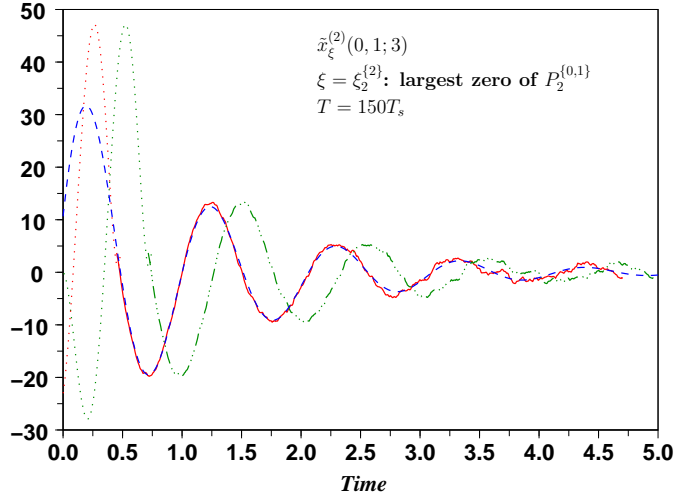


Fig. 6 Non minimal, delayed second order derivative estimation.

for $q = 1$. This stems from the identity

$$h_{\kappa,\mu}^{MV}(\tau) = \frac{\kappa + \mu + 2n + 2}{2(\mu + n + 1)} h_{\kappa,\mu}^{Min}(\tau) + \frac{m\mu - \kappa}{2(\kappa + \mu + 2n + 2)} h_{\kappa+1,\mu}^{Min}(\tau)$$

where the superscripts “MV” and “Min” refer to the “Minimum Variance” and “Minimal” estimators respectively. An illustration is given in the figure 7 below, with the settings of the two preceding simulation.

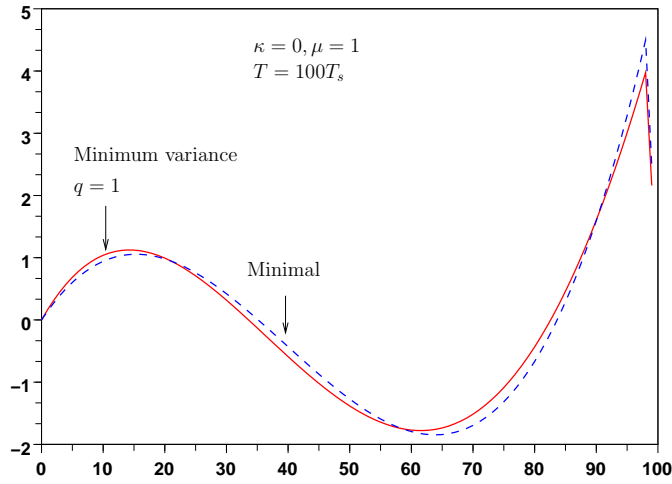


Fig. 7 Minimum variance vs Minimal second order derivative filters.

Simulation	Fig. 2	Fig. 3	Fig. 4	Fig. 5	Fig. 6	Fig. 7
SBR in dB	23.2	12	33.7	18.1	28.3	17.9
SNRout in dB	26.5	12	26.6	22.6	24.1	22.9

Table 1 Bias and variance

Finally, let us notice that all the preceding simulation results are based on one sample realization of the noisy signal. Now, in order to validate the presented results, we investigate the bias and variance in each simulation above. Recall that the signal derivatives estimates are obtained as output of FIR filters with a noisy input. So, with the zero mean white noise assumption, the associated output noise variance is readily computed by taking the ℓ_2 sequence norm of the corresponding impulse response, $\{c_m\}$, multiplied by the variance of the input noise. The output signal to noise ratio (SNRout) is therefore given by

$$\text{SNRout} = \frac{\frac{1}{L-M} \sum_{\ell=M}^L |x_\ell^{(n)}|^2}{\sigma_\varpi^2 \sum_{m=0}^{M-1} |c_m|^2}.$$

Concerning the bias, we use 100 Monte-Carlo simulation runs to compute the sample means $\mathbb{E}_s\{\tilde{x}^{(n)}(\ell)\}$. The relative sample mean squared error, due to the mismodelling (the bias term) and to the numerical approximation of the integrals, will be called the *signal to bias ratio*. It is computed as:

$$\text{SBR} = \frac{\sum_{\ell=M}^L |x_\ell^{(n)}|^2}{\sum_{\ell=M}^L |x_\ell^{(n)} - \mathbb{E}_s\{\tilde{x}^{(n)}(\ell)\}|^2}.$$

Table 4.2 displays the signal to bias ratio (SBR) and the output signal to noise ratio (SNRout). The values are expressed in decibel (dB). Here, $L = 1000$ is the signal duration, that is: $LT_s = 5$. Each column, except the last one, represents the results obtained in the context of the simulation of the figure referenced in the first row of the table. The last column displays the results obtained with the second order minimum variance derivative estimator filter show in figure 7.

Proofs

Proof (Proof of Lemma 2) A direct inspection of (10) shows that the maximum power of s , appearing in the coefficients of the differential operator $\Pi_{\kappa,\mu}^{N,n}$, is $s^{-\mu-1}$. That $\Pi_{\kappa,\mu}^{N,n}$ is in finite-integral form is thus clear since $\mu \geq 0$.

Next apply this differential operator to \hat{x}_N . The $(N-n)^{th}$ order derivative of the right hand side of (4) results in a polynomial of degree n , the constant term of which is equal to $(N-n)!x^{(n)}(0)$. If we divide this polynomial by s and take the $(n+\kappa)^{th}$ order derivative of the result, then we immediately see that there is only one remaining term which, when divided by $s^{N+\mu+1}$, yields $\varphi_{\kappa,\mu}(s)x^{(n)}(0)$, where $\varphi_{\kappa,\mu}(s)$ is given by (11). \square

Proof (Proof of Theorem 1) Set $q = N - n$ and $p = n + \kappa$. The first part follows by direct inspection, upon writing $D = \frac{d^p}{ds^p} \frac{1}{s} \frac{d^q}{ds^q} s^{q+1} s^n$ in the form

$$\begin{aligned} D &= \sum_{i=0}^q \binom{q}{i} \frac{(q+1)!}{(q+1-i)!} \frac{d^p}{ds^p} s^{q-i} \frac{d^{q-i}}{ds^{q-i}} s^n \\ &= \sum_{i=0}^q \sum_{j=i}^{\min(p+i, q)} a_{i,j} s^{q-j} \frac{d^{p+q-j}}{ds^{p+q-j}} s^n, \end{aligned} \quad (50)$$

where

$$a_{i,j} = \binom{q}{i} \binom{p}{j-i} \frac{(q+1)!}{(q+1-i)(q-j)!}.$$

Dividing D by $s^{\mu+N+1}$ and rearranging the terms in the summation above allows one to get:

$$\Pi_{\kappa, \mu}^{N, n} = \sum_{j=0}^q \left(\sum_{i=m(j)}^j a_{i,j} \right) \Pi_{\kappa+q-j, \mu+j}^n,$$

where, if $q \leq p$ then $m(j) \equiv 0$ and else $m(j) = \max(0, j - p)$. Using the expression (11) and the relation (12), we get the linear combination in (14) with coefficients given by:

$$\lambda_j = (-1)^{q-j} \frac{(n + \kappa + q - j)!}{p!q!} \sum_{i=m(j)}^j a_{i,j}.$$

It now remains to show that, (12) is a \mathbb{Q} -affine combination. That the coefficients λ_j are rational numbers is indeed plain. By rearranging the factors of the numerator and of denominator, we obtain the simpler expression:

$$\begin{aligned} \lambda_j &= (-1)^{q-j} \binom{p+q-j}{p} \sum_{i=0}^j \binom{p}{j-i} \binom{q+1}{i} \\ &= (-1)^{q-j} \binom{p+q+1}{p} \binom{p+q+1}{j} \end{aligned} \quad (51)$$

for $q \leq p$. Define $I(q) = \sum_{j=0}^q \lambda_j$. Then, with the help of the identity $\binom{m+1}{\ell} = \binom{m}{\ell} + \binom{m}{\ell-1}$, $\ell > 0$, we have the recurrence relation

$$\begin{aligned} I(q+1) &= I(q) + \sum_{j=0}^{q+1} (-1)^j \binom{p+j}{j} \binom{p+q+1}{p+j} \\ &= I(q) + \binom{p+q+1}{p} \sum_{j=0}^{q+1} (-1)^j \binom{q+1}{j} = I(q), \end{aligned}$$

showing that $I(q)$ is constant. Therefore, we get $\sum_{j=0}^q \lambda_j = 1$ upon noting that $I(1) = 1$. \square

Proof (Proof of Proposition 1) Observe that $x_{LS,1}^{(n)}(Tt)$ is defined by

$$x_{LS,1}^{(n)}(Tt) = \frac{\langle P_0^{\{\kappa, \mu\}}(\tau), x^{(n)}(T\tau) \rangle}{\|P_0^{\{\kappa, \mu\}}\|^2} + \frac{\langle P_1^{\{\kappa, \mu\}}(\tau), x^{(n)}(T\tau) \rangle}{\|P_1^{\{\kappa, \mu\}}\|^2} P_1^{\{\kappa, \mu\}}(t).$$

For $t = \xi_1$ as in (22), i.e., the zero of $P_1^{\{\kappa, \mu\}}(t)$, the second term in the right hand side of the equality above vanishes. Hence, we recognize the contribution due to the signal in equation (17). The contribution due to the noise is denoted by the term $e_{\varpi}(0; \kappa, \mu)$ in (21). \square

Proof (Proof of Theorem 2) To simplify the expressions to come, we shall write $\tilde{x}^{(n)}(\kappa, \mu; N)$ in place of $\tilde{x}^{(n)}(0; \kappa, \mu; N)$. Begin with the following lemma.

Lemma 3 *The recurrence relation*

$$\tilde{x}^{(n)}(\kappa, \mu; N) = \lambda \tilde{x}^{(n)}(\kappa, \mu + 1; N - 1) + (1 - \lambda) \tilde{x}^{(n)}(\kappa + 1, \mu; N - 1), \quad (52)$$

where λ is defined by $\lambda = \frac{N+\kappa+1}{N-n}$, holds for any κ, μ, n, N positive, with $N > n$.

Proof For the sake of simplicity, we ignore the noise. This is without any loss of generality in the present proof. Set $q = N - n$, $\nu = \mu + N + 1$ and

$$A_N(\kappa, \mu) = (-1)^{\kappa+n} \frac{(\kappa+n)!(\mu+n)!}{(\kappa+\mu+N+n+1)!} \tilde{x}^{(n)}(\kappa, \mu; N). \quad (53)$$

Also, define $x_T(t) \triangleq x(Tt)$. A direct application of the rules of operational calculus shows that:

$$A_N(\kappa, \mu) = \int_0^1 (-1)^{\kappa+n} t^{\kappa+n} \frac{(1-t)^{\nu-1}}{(\nu-1)!} \int_0^t (-1)^q \tau^q x_T^{(N+1)}(\tau) d\tau dt. \quad (54)$$

Integrate by parts for the second integral. Then, we may rewrite the above expression as:

$$A_N(\kappa, \mu) = q A_{N-1}(\kappa, \mu+1) + \frac{(-1)^{\kappa+N}}{(\mu+N)!} \int_0^1 t^{\kappa+n+q} (1-t)^{\mu+n+q} x_T^{(N)}(\tau) dt. \quad (55)$$

Let $B_{N-1}(\kappa, \mu)$ be defined by the second term of the right hand side of the previous equation: $B_{N-1}(\kappa, \mu) \triangleq A_N(\kappa, \mu) - q A_{N-1}(\kappa, \mu+1)$. Now reconsider (54), and integrate by parts again but this time, with respect to the first integral. This will result in:

$$A_N(\kappa, \mu) = -\frac{1}{\kappa+n+1} \{A_N(\kappa+1, \mu-1) - B_N(\kappa, \mu-1)\}. \quad (56)$$

Combining (55) and (56), we arrive at the recurrence relation:

$$A_N(\kappa, \mu) = (\kappa+N+1)A_{N-1}(\kappa, \mu+1) + A_{N-1}(\kappa+1, \mu). \quad (57)$$

Finally, the relation (52) follows by considering the definition (53). \square

We proceed by induction.

For $q = 1$, equation (52) coincides with (14), with $\lambda_1 = n + \kappa + 2 = \lambda$ and $\lambda_0 = 1 - \lambda$. Using (17), it can be written as

$$\begin{aligned} \tilde{x}^{(n)}(\kappa, \mu; n+1) = \int_0^1 \{ & \lambda \gamma_{\kappa, \mu+1, n} w_{\{\kappa, \mu+1\}}(\tau) \\ & + (1-\lambda) \gamma_{\kappa+1, \mu, n} w_{\{\kappa+1, \mu\}}(\tau) \} y^{(n)}(T\tau) d\tau. \end{aligned} \quad (58)$$

Set $\alpha = \mu + n$ and $\beta = \kappa + n$.

Now, it is immediate to verify that the bracketed term under the integral above has the form $g(\tau) w_{\{\kappa, \mu\}}(\tau)$ where g is a polynomial of degree 1. The following identity

$$\eta_m(\kappa, \mu) \triangleq \frac{P_m^{\{\kappa, \mu\}}(0)}{\|P_m^{\{\kappa, \mu\}}\|^2} = \frac{(-1)^m \Gamma(2m + \alpha + \beta + 2)}{m! \Gamma(\beta + 1) \Gamma(m + \alpha + 1)}$$

may be verified directly from [1]. Then it is a matter of straightforward algebra to check that the polynomial g expresses as:

$$g(\tau) = \frac{1}{\|P_0^{\{\kappa, \mu\}}\|^2} + \eta_1(\kappa, \mu) P_1^{\{\kappa, \mu\}}(\tau),$$

where we recall that $P_1^{\{\kappa, \mu\}}(\tau) = \tau - \xi_1$, with ξ_1 given by (22). Plugging this expression of g into (58) immediately shows that (28) is valid for $q = 1$.

Suppose (28) holds for a given $q > 1$. Then, we obtain from (52),

$$\begin{aligned} \tilde{x}^{(n)}(\kappa, \mu; n+q+1) = \lambda \sum_{m=0}^q \langle P_m^{\{\kappa, \mu+1\}}, x^{(n)} \rangle_{\{\kappa, \mu+1\}} \eta_m(\kappa, \mu+1) \\ + (1-\lambda) \sum_{m=0}^q \langle P_m^{\{\kappa+1, \mu\}}, x^{(n)} \rangle_{\{\kappa+1, \mu\}} \eta_m(\kappa+1, \mu). \end{aligned} \quad (59)$$

The following miscellaneous recurrence relations, which are adapted from [1, p.782], will be useful:

$$(1-t)P_m^{\{\kappa, \mu+1\}}(t) = a_m(\kappa, \mu)P_m^{\{\kappa, \mu\}}(t) - P_{m+1}^{\{\kappa, \mu\}}(t) \quad (60)$$

$$tP_m^{\{\kappa+1, \mu\}}(t) = a_m(\mu, \kappa)P_m^{\{\kappa, \mu\}}(t) + P_{m+1}^{\{\kappa, \mu\}}(t) \quad (61)$$

where

$$a_m(\kappa, \mu) = -\frac{P_m^{\{\kappa, \mu\}}(0)}{P_{m-1}^{\{\kappa, \mu\}}(0)} = \frac{(m+\alpha+1)(m+\alpha+\beta+1)}{(2m+\alpha+\beta+1)(2m+\alpha+\beta+2)} \quad (62)$$

With these relations, the equation (59) becomes

$$\tilde{x}^{(n)}(\kappa, \mu; n+q+1) = \sum_{m=0}^{q+1} \zeta_m \langle P_m^{\{\kappa, \mu\}}, x^{(n)} \rangle, \quad (63)$$

with

$$\begin{aligned} \zeta_0 &= \lambda a_0(\kappa, \mu) \eta_0(\kappa, \mu+1) + (1-\lambda) a_0(\mu, \kappa) \eta_0(\kappa+1, \mu), \\ \zeta_{q+1} &= -\lambda \eta_q(\kappa, \mu+1) + (1-\lambda) \eta_q(\kappa+1, \mu), \\ \zeta_m &= \lambda [a_m(\kappa, \mu) \eta_m(\kappa, \mu+1) - \eta_{m-1}(\kappa, \mu+1)] \\ &\quad + (1-\lambda) [a_m(\mu, \kappa) \eta_m(\kappa+1, \mu) - \eta_{m-1}(\kappa+1, \mu)] \\ &\quad 1 \leq m \leq q, \end{aligned}$$

Now, it is straightforward to verify that the following identities hold:

$$\begin{aligned} \zeta_0 &= \frac{1}{\|P_0^{\{\kappa, \mu\}}\|^2}, \\ \zeta_{q+1} &= \eta_{q+1}(\kappa, \mu), \\ \eta_m(\kappa, \mu) &= a_m(\kappa, \mu) \eta_m(\kappa, \mu+1) - \eta_{m-1}(\kappa, \mu+1) \\ &= a_m(\mu, \kappa) \eta_m(\kappa+1, \mu) - \eta_{m-1}(\kappa+1, \mu) \\ &= \zeta_m, \quad m = 1, \dots, q \end{aligned}$$

This shows that if the relation (28) is valid for $q > 1$, then it must also be valid for $q+1$. Since it has been established for $q=1$, we conclude that (28) holds for all $q \geq 1$. \square

Proof (Proof of Proposition 2) As for the equation (58) in the previous proof, it is easy to see that the linear combination $\sum_{\ell=0}^q \lambda_\ell \tilde{x}^{(n)}(0; \kappa_\ell, \mu_\ell)$ is given by

$$\sum_{\ell=0}^q \lambda_\ell \tilde{x}^{(n)}(0; \kappa_\ell, \mu_\ell) = \int_0^1 B(\tau) w_{\{\kappa, \mu\}}(\tau) y^{(n)}(T\tau) d\tau = \langle B(\tau), y^{(n)}(T\tau) \rangle. \quad (64)$$

with

$$B(\tau) = \sum_{\ell=0}^q \lambda_\ell \gamma_{\kappa_\ell, \mu_\ell, n} \tau^{q-\ell} (1-\tau)^\ell = \sum_{\ell=0}^q \lambda_\ell \phi_\ell b_{\ell, q}(\tau), \quad (65)$$

where

$$b_{\ell, q}(\tau) \triangleq \binom{q}{\ell} \tau^{q-\ell} (1-\tau)^\ell, \quad \ell = 0, \dots, q,$$

are the Bernstein basis polynomials [23] of degree q and $\phi_\ell \triangleq \frac{\gamma_{\kappa_\ell, \mu_\ell, n}}{\binom{q}{\ell}}$.

Since the set $\{b_{\ell, q}(\tau), \ell = 0, \dots, q\}$ also forms a basis for the subspace \mathcal{H}_q (see (25)), we may equate

$$\mathcal{K}_q(\xi, \tau) = \sum_{\ell=0}^q \sum_{m=0}^q b_{m, q}(\xi) (\mathcal{B}^{-1})_{m, \ell} b_{\ell, q}(\tau). \quad (66)$$

Here, the matrix \mathcal{B} stands for the Gramian of the Bernstein polynomials $b_{\ell,q}(\cdot)$, with respect to the inner product of \mathcal{H}_q .

Now, observe that the first part of the proposition claims that, for each $\xi \in [0, 1]$, the equation

$$B(\tau) = \mathcal{K}_q(\xi, \tau) \quad (67)$$

admits a unique solution for $(\lambda_0, \dots, \lambda_q)$. With the notations

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_0 \\ \vdots \\ \lambda_q \end{bmatrix}; \quad \mathbf{b}_q(\tau) = \begin{bmatrix} b_{0,q}(\tau) \\ \vdots \\ b_{q,q}(\tau) \end{bmatrix}; \quad \Phi = \begin{bmatrix} \phi_0 & 0 \\ & \ddots \\ 0 & \phi_q \end{bmatrix},$$

and with (65) and (66), this equation becomes equivalent to $\Phi \boldsymbol{\lambda} = \mathcal{B}^{-1} \mathbf{b}_q(\xi)$. It is clear that it admits the unique solution:

$$\boldsymbol{\lambda}(\xi) = \Phi^{-1} \mathcal{B}^{-1} \mathbf{b}_q(\xi). \quad (68)$$

It now remains to establish the second part of the proposition. To proceed, observe that the entries of the Gramian matrix \mathcal{B} are readily identified as:

$$(\mathcal{B})_{i,j} = \binom{q}{i} \left\| P_0^{\{\kappa+2q-(i+j), \mu+(i+j)\}} \right\|^2 \binom{q}{j}.$$

Next, using (20) and some straightforward algebra allows one to verify that

$$(\mathcal{B})_{i,j} \phi_j = \frac{\binom{\mu+n+i+j}{\mu+n+j} \binom{\kappa+n+2q-(i+j)}{\kappa+n+q-j}}{\binom{\mu+\kappa+2n+2q+1}{q}}.$$

Now, with the help of the classical identity

$$\sum_m \binom{m}{\zeta} \binom{i-m}{j-\zeta} = \binom{i+1}{j+1},$$

which holds independently of ζ , one may readily check that

$$\sum_{j=0}^q (\mathcal{B})_{i,j} \phi_j = 1, \quad \forall i.$$

We have therefore shown that $\mathbf{u}^t = [1 \ 1 \ \dots \ 1]$ is a left eigenvector of the matrix $\mathcal{B}\Phi$, and the associated eigenvalue is 1. Last, since the Bernstein polynomials $b_{\ell,q}(\tau)$, $\ell = 0, \dots, q$ form a partition of unity [23], we get:

$$\sum_{\ell=0}^q \lambda_\ell(\xi) = \mathbf{u}^t \boldsymbol{\lambda}(\xi) = \mathbf{u}^t \{\mathcal{B}\Phi\}^{-1} \mathbf{b}_q(\xi) = \mathbf{u}^t \mathbf{b}_q(\xi) \equiv 1.$$

Finally, note that since the Bernstein basis polynomials are strictly positive in $(0, 1)$, the definition of $B(\cdot)$ in (65) and the equation (67) imply that the condition $\lambda_\ell(\xi) \geq 0$ for all ℓ is equivalent to

$$\left\| P_q^{\{\kappa, \mu\}} \right\|^2 \mathcal{K}_q(\xi, \tau) = \frac{P_{q+1}^{\{\kappa, \mu\}}(\tau) P_q^{\{\kappa, \mu\}}(\xi) - P_q^{\{\kappa, \mu\}}(\tau) P_{q+1}^{\{\kappa, \mu\}}(\xi)}{\tau - \xi} > 0,$$

for all $\tau \in (0, 1)$. Now, it is clear that this cannot hold. The above equality is the classical Christoffel-Darboux formula.

□

Proof (Proof of Proposition 3) From the first order approximations of $|e_{R_N}(t; \kappa, \mu; n+q)|$ as in (38) and of $|e_{R_{N-1}}^{\xi_1^{\{q\}}}(t; \kappa, \mu; n+q-1)|$, we may write

$$\frac{|e_{R_{N-1}}^{\xi_1^{\{q\}}}(t; \kappa, \mu; n+q-1)|}{|e_{R_N}(t; \kappa, \mu; n+q)|} \approx \left| \frac{P_{q+1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})}{P_{q+1}^{\{\kappa, \mu\}}(0)} \right|.$$

We now proceed to bound the right hand side of the above expression. For that, recall first the three terms recurrence relations generating the family of the Jacobi polynomials in $[0, 1]$, associated with the weight defined in (18):

$$P_0^{\{\kappa, \mu\}}(t) = 1, \quad P_1^{\{\kappa, \mu\}}(t) = t - \frac{\kappa + n + 1}{\mu + \kappa + 2n + 2} \quad (69)$$

$$P_{m+1}^{\{\kappa, \mu\}}(t) = (t - c_m)P_m^{\{\kappa, \mu\}}(t) - d_m P_{m-1}^{\{\kappa, \mu\}}(t) \quad (70)$$

where

$$d_m = \frac{\|P_m^{\{\kappa, \mu\}}\|^2}{\|P_{m-1}^{\{\kappa, \mu\}}\|^2} = \frac{m(m+\beta)(m+\alpha+\beta)(m+\alpha)}{(2m+\alpha+\beta-1)(2m+\alpha+\beta)^2(2m+\alpha+\beta+1)},$$

with $\alpha = \mu + n$ and $\beta = \kappa + n$.

For $t = \xi_1^{\{q\}}$, we get $P_{q+1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}}) = -d_q P_{q-1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})$ so that we also have:

$$d_q = -P_{q+1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})/P_{q-1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}}).$$

On the other hand, setting $t = 0$ in (61) provides us with the recurrence

$$\begin{aligned} P_{q+1}^{\{\kappa, \mu\}}(0) &= -a_q(\mu, \kappa)P_q^{\{\kappa, \mu\}}(0) \\ &= a_q(\mu, \kappa)a_{q-1}(\mu, \kappa)P_{q-1}^{\{\kappa, \mu\}}(0). \end{aligned}$$

Using all these relations and owing that the polynomials $P_i^{\{\kappa, \mu\}}$ are monic, we get:

$$\left| \frac{P_{q+1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})}{P_{q+1}^{\{\kappa, \mu\}}(0)} \right| = \left| \frac{P_{q+1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})}{P_{q-1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})} \right| \cdot \left| \frac{P_{q-1}^{\{\kappa, \mu\}}(0)}{P_{q+1}^{\{\kappa, \mu\}}(0)} \right| \cdot \left| \frac{P_{q-1}^{\{\kappa, \mu\}}(\xi_1^{\{q\}})}{P_{q-1}^{\{\kappa, \mu\}}(0)} \right| \quad (71)$$

$$= d_q \cdot \frac{1}{a_q(\mu, \kappa)a_{q-1}(\mu, \kappa)} \cdot \prod_{j=1}^{q-1} \left(1 - \frac{\xi_1^{\{q\}}}{\xi_j^{\{q-1\}}} \right) \quad (72)$$

Defining

$$\theta_q(\kappa, \mu) \triangleq d_q \cdot \frac{1}{a_q(\mu, \kappa)a_{q-1}(\mu, \kappa)}$$

we end up with the expression (40) by a direct computation. Also, it is readily checked that $\theta_q(\kappa, \mu) < 1$. On the other hand, notice that the interlacing property $0 < \xi_1^{\{q\}} < \xi_1^{\{q-1\}} < \xi_j^{\{q-1\}}$, $j = 2, \dots, q-1$ implies that each term $1 - \frac{\xi_1^{\{q\}}}{\xi_j^{\{q-1\}}}$ of the product in (72) is positive and strictly less than 1. This product is therefore much lower than one and particular, we have

$$\prod_{j=1}^{q-1} \left(1 - \frac{\xi_1^{\{q\}}}{\xi_j^{\{q-1\}}} \right) < \left(1 - \frac{z_q(\kappa, \mu)}{Z_{q-1}(\kappa, \mu)} \right)^{q-1},$$

$z_q(\kappa, \mu)$ and $Z_q(\kappa, \mu)$ the lower and upper bounds of the smallest and largest zeros of $P_q^{\{\kappa, \mu\}}$, respectively. \square

Acknowledgements The authors are grateful to the referees for the comments and suggestions which improve the clarity of this paper.

References

1. Abramowitz, M., Stegun, I.A. (eds.): Handbook of mathematical functions. Dover Publications (1965)
2. Al-Alaoui, M.A.: A class of second-order integrators and low-pass differentiators. *IEEE Trans. Circuits Syst. I* **42**(4), 220–223 (1995)
3. Alpay, D.: Algorithme de Schur, espaces à noyau reproduisant et théorie des systèmes, *Panoramas et Synthèses*, vol. 6. Société mathématique de France (1998)
4. Aronszajn, N.: Theory of reproducing kernels. *Trans. AMS* **68**(3), 337–404 (1950)
5. Chen, C.K., Lee, J.H.: Design of high-order digital differentiators using L_1 error criteria. *IEEE Trans. Circuits Syst. II* **42**(4), 287–291 (1995)
6. Chitour, Y.: Time-varying high-gain observers for numerical differentiation. *IEEE Trans. Automat. Contr.* **47**, 1565–1569 (2002)
7. Dabroom, A.M., Khalil, H.K.: Discrete-time implementation of high-gain observers for numerical differentiation. *International Journal of Control* **72**, 1523–1537 (1999)
8. Diop, S., Grizzle, J.W., Chaplais, F.: On numerical differentiation algorithms for non-linear estimation. In: *Proc. CDC. Sydney* (2000)
9. Duncan, T.E., Mandl, P., Pasik-Duncan, B.: Numerical differentiation and parameter estimation in higher-order linear stochastic systems. *IEEE Trans. Automat. Contr.* **41**, 522–532 (1996)
10. Fliess, M., Join, C., Sira Ramírez, H.: Closed-loop fault-tolerant control for uncertain nonlinear systems. In: T. Meurer, K. Graiche, E. Gilles (eds.) *Control and Observer Design for Nonlinear Finite and Infinite Dimensional Systems, Lect. Notes Control Informat. Sci.*, vol. 322, pp. 217–233. Springer (2005)
11. Fliess, M., Join, C., Mboup, M., M., H.S.R.: Compression différentielle de transitoires bruités. *CRAS, Série 1, Mathématiques* **339**, 821–826 (2004)
12. Fliess, M., Join, C., Mboup, M., Sira Ramírez, H.: Analyse et représentation de signaux transitoires : application à la compression, au débruitage et à la détection de ruptures. In: *Actes 20^e Coll. GRETSI. Louvain-la-Neuve* (2005). URL <http://hal.inria.fr/inria-00001115>
13. Fliess, M., Join, C., Sedoglavic, A.: Estimation des dérivées d'un signal multidimensionnel avec applications aux images et aux vidéos. In: *Actes 20^e Coll. GRETSI. Louvain-la-Neuve* (2005). URL <http://hal.inria.fr/inria-00001116>
14. Fliess, M., Mboup, M., Mounier, H., Sira-Ramírez, H.: Questioning some paradigms of signal processing via concrete examples. In: H. Sira-Ramírez, G. Silva-Navarro (eds.) *Algebraic Methods in Flatness, Signal Processing and State Estimation*. Innovación ed. Lagares, México (2003)
15. Fliess, M., Sira-Ramírez, H.: An algebraic framework for linear identification. In: *ESAIM: COCV*, vol. 9. SMAI (2003). URL: <http://www.emath.fr/cocv>
16. Fliess, M., Sira Ramírez, H.: Control via state estimations of some nonlinear systems. In: *Proc. Symp. Nonlinear Control Systems (NOLCOS'04)*. Stuttgart (2004). URL <http://hal.inria.fr/inria-00001096>
17. Haykin, S., Van Veen, B.: *Signals and Systems*, 2nd edn. John Wiley & Sons (2002)
18. Ibrir, S.: Online exact differentiation and notion of asymptotic algebraic observers. *IEEE Trans. Automat. Contr.* **48**, 2055–2060 (2003)
19. Ibrir, S.: Linear time-derivatives trackers. *Automatica* **40**, 397–405 (2004)
20. Ibrir, S., Diop, S.: A numerical procedure for filtering and efficient high-order signal differentiation. *Int. J. Appl. Math. Compt. Sci.* **14**, 201–208 (2004)
21. Ismail, M.E.H., Li, X.: Bound on the extreme zeros of orthogonal polynomials. *Proceedings of the AMS* **115**(1), 131–140 (1992)
22. Levant, A.: Higher-order sliding modes, differentiation and output-feedback control. *International Journal of Control* **76**, 924–941 (2003)
23. Lorentz, G.G.: *Bernstein Polynomials*, 2nd edn. AMS Chelsea Publishing (1986)
24. Massart, P.: Concentration inequalities and model selection, *Lecture Notes in Mathematics*, vol. 1896. Springer, Berlin (2007)
25. Mboup, M.: Parameter estimation via differential algebra and operational calculus. *Tech. rep., Submitted to Signal Processing* (2007)

-
26. Mboup, M., Join, C., Fliess, M.: A revised look at numerical differentiation with an application to nonlinear feedback control. In: 15th Mediterranean conference on Control and automation (MED'07). Athenes, Greece (2007)
 27. Mikusiński, J.: Operational Calculus, vol. 1. PWN Varsovie & Oxford University Press, Oxford (1983)
 28. Mikusiński, J., Boehme, T.K.: Operational Calculus, vol. 2. PWN Varsovie & Oxford University Press, Oxford (1987)
 29. Rader, C.M., Jackson, L.B.: Approximating noncausal IIR digital filters having arbitrary poles, including new Hilbert transformer designs, via forward/backward block recursion. *IEEE Trans. Circuits Syst. I* **53**(12), 2779–2787 (2006)
 30. Richard, J.: Time delay systems: An overview of some recent advances and open problems. *Automatica* **10**, 1667–1694 (2003)
 31. Roberts, R.A., Mullis, C.T.: Digital signal processing. Addison-Wesley (1987)
 32. Saitoh, S.: Theory of reproducing kernels and its applications. Pitman Research Notes in Mathematics. Longman Scientific & Technical, UK (1988)
 33. Seuret, A., Dambrine, M., Richard, J.: Robust exponential stabilization for systems with time-varying delays. In: 5th IFAC Workshop on Time Delay Systems. Leuven, Belgium (2004)
 34. Su, Y.X., Zheng, C.H., Mueller, P.C., Duan, B.Y.: A simple improved velocity estimation for low-speed regions based on position measurements only. *IEEE Trans. Control Syst. Technology* **14**, 937–942 (2006)
 35. Szegő, G.: Orthogonal polynomials, 3rd edn. AMS, Providence, RI (1967)
 36. Yosida, K.: Operational Calculus - A Theory of Hyperfunctions. Springer, New York (1984)